

Durham Research Online

Deposited in DRO:

24 November 2016

Version of attached file:

Accepted Version

Peer-review status of attached file:

Peer-reviewed

Citation for published item:

Merrell, C. (2017) 'Understanding monitoring in the United Kingdom context.', in Monitoring the quality of education in schools : examples of feedback into systems from developed and emerging economies. Rotterdam/Boston/Taipei: Sense Publishers, pp. 93-106.

Further information on publisher's website:

<https://www.sensepublishers.com/catalogs/bookseries/other-books/monitoring-the-quality-of-education-in-schools/>

Publisher's copyright statement:

Additional information:

Use policy

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a [link](#) is made to the metadata record in DRO
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the [full DRO policy](#) for further details.

**CHRISTINE MERRELL, DIRECTOR OF RESEARCH AND
READER IN EDUCATION**

UNDERSTANDING MONITORING IN THE UNITED KINGDOM CONTEXT

Centre for Evaluation & Monitoring, Durham University, England

This chapter provides an overview of the recent and current policies for assessing and monitoring pupils' academic progress across the four countries of the UK; England, Northern Ireland, Scotland and Wales. It gives examples of the uses of the data at pupil, school and system levels and discusses some of the issues associated with these uses. Finally, a family of non-statutory large-scale monitoring systems, established by the Centre for Evaluation & Monitoring (CEM) at Durham University, England, and taken up by many schools as an alternative to nationally mandated schemes are presented. The uses of the information from CEM's systems, in contrast to the data from statutory systems, are described.

OVERVIEW OF THE OFFICIAL SYSTEMS FOR ASSESSING & MONITORING PUPILS' PROGRESS IN THE UNITED KINGDOM

Over a period of decades, the powers to set educational policy and provision have been gradually devolved from the central United Kingdom government in London to assemblies in Northern Ireland and Wales, and the Scottish Parliament, and following this process of devolution four different educational systems have evolved. The development of each of the systems and some of the issues associated with that development are described in more detail in the next part of the chapter.

The English System

In England, the Educational Reform Act of 1988 marked significant changes to the education system. The Act was intended to lead to a rise in educational standards and as part of the reform, a National Curriculum was introduced and, with it, a statutory assessment framework. The National Curriculum was intended to provide a broad, balanced and coherent educational experience for children aged between 5 and 16. Key Stages of education were introduced; Key Stage 1 for ages 5 – 7; Key Stage 2 for ages 7 – 11, Key Stage 3 for ages 11 – 14 and Key Stage 4 for ages 14 – 16. The end of Key Stage 4 was the end of compulsory education at which point pupils sat examinations set by awarding bodies. More recently, the

CHRISTINE MERRELL

Early Years Foundation Stage was introduced for children from birth up to the age of 5.

Statutory tests in literacy, mathematics and science, conducted at the end of Key Stages 1, 2 and 3, were introduced. The first of the new statutory tests to be used was for the end of Key Stage 1 and were first taken in 1991. The implementation of the tests was turbulent. Teachers' unions frequently called for boycotts. Assessment at the end of Key Stage 1 quickly shifted from tests and tasks to teacher assessment. At Key Stage 3, there were problems with the development and piloting of the tests in 1993 and in that year they were not used by the majority of schools (Whetton, 2009). Problems continued, especially with the external marking of the test papers and in October 2008, the government announced that the end of Key Stage 3 tests would be replaced with teacher assessment. The end of Key Stage 2 science tests were dropped in 2009. In 2010, around 7,000 of England's 17,000 primary schools boycotted the end of Key Stage 2 tests and, following problems with external marking, the writing test was changed from being externally marked to being marked by teachers in school.

In 2011, the English government commissioned an independent review of testing arrangements at the end of Key Stage 2. Evidence from many sources was gathered, including expert opinion, and one of the resulting recommendations was to increase the amount of assessment based on teacher judgement rather than tests (Bew, 2011).

The original intention of the tests was to provide formative and diagnostic information to guide teachers' practice, to provide summative information about the levels of attainment reached and to provide evaluative information by aggregating results to class and school level to indicate the functioning of the curriculum, teachers and schools (Task Group on Assessment and Testing, TGAT, 1988). Although the initial recommendations suggested a broad range of uses for the data, the main focus rapidly evolved towards accountability. The test scores from all pupils were centrally collated and the percentage of pupils achieving the expected level of attainment or higher was made publicly available. In 1992, national newspapers first printed league tables of the schools with the highest and lowest scores. Schools with results that were below expectations were to be held to account, which led to teachers spending months preparing their pupils for the tests. 'Booster classes' were set up in schools to provide extra support and tuition in the months leading up to the test for those children who were on the borderline of achieving the expected level of attainment to try and ensure that they would make the grade on the day of the test. Much class time was devoted to preparing for the tests in the months leading up to their administration and from this grew fears of a narrowing of the curriculum. The Cambridge Primary Review, which was a significant report on many aspects of primary education in England, reported evidence that national tests and league tables were indeed linked to a narrowing of the curriculum, limiting children's learning (Alexander, 2010). Ambitious targets for pupil attainment were set and schools that failed to achieve them were in the uncomfortable position of having to explain themselves to inspectors, the local education authority and the public. In small schools where, perhaps, one or two

pupils did not attain their expected level on the day of the test, their under-performance could have a significant effect on the average score. Parents too were caught up in the ambition for their children to perform at the expected level and commercial publishers produced books containing practice questions for parents to buy to coach their children.

There is some evidence to suggest that the introduction of the statutory end of Key Stage tests had an adverse effect of pupils' self-esteem and caused stress. Davis and Brember (1998 and 1999) studied the self-esteem of pupils in five schools over a period of eight years, beginning two years before the introduction of the statutory tests. They saw a decline in pupils' self-esteem at the end of Key Stage 1 for the first four years of the study, with the greatest decline coinciding with the introduction of the tests. In the early days of their introduction, the administration procedure for the tests was complicated (James, 2013). There was an improvement in the self-esteem of the end of Key Stage 1 cohort which coincided with a simplification in the administration procedure of the tests and teachers becoming more accustomed to them. In a second study, Reay and Wiliam (1999) used a mixture of focus groups, individual interviews and classroom observations to investigate the views of a class of pupils aged 11 years towards the statutory end of Key Stage 2 test. They noted considerable changes in the pupils during the term leading up to the test when they expressed an awareness of the consequences of the statutory assessment and anxiety about failure. A Briefing Paper published by the National Union of Teachers (2006) brought together evidence from previously published research and the results of a survey that pointed to an association between statutory end of Key Stage tests and an increase in the stress and anxiety of pupils. The formative use of the statutory end of Key Stage tests became the last thing on everyone's minds.

The publication of the percentage of pupils achieving at or higher than the expected level of attainment presented a narrow view of school performance and did not take into account how much progress pupils had made between Key Stages. There was a further problem; each pupil received a score from the end of Key Stage test, and these test scores were converted to National Curriculum levels for reporting purposes. There are eight National Curriculum levels which span Key Stages 1 – 3 and are criterion referenced against the curriculum, therefore are stable over time. Children are expected to reach level 2 by the end of Key Stage 1 and Level 4 by the end of Key Stage 2. These levels represent a broad range of attainment and yet the difference of just one mark on the end of Key Stage test, which would be expected given that all tests have an error of measurement, if a child re-took the test, could mean that s/he was assigned a different National Curriculum level. William (2000) estimated the proportion of children who were potentially misclassified at the end of Key Stage 2. That is the proportion of pupils whose range of possible test scores overlapped with the cut-off points for National Curriculum levels and so with a slightly different score on a different day would be assigned a different National Curriculum level. For a test with a reliability of 0.85, which is close to the end of Key Stage 2 science test in 2007, William estimated that 27% of pupils were misclassified. As the reliability of a test improves, the

proportion of misclassified pupils decreases but only slightly, and for a reliability of 0.9, which was similar to the reading, spelling and mental mathematics tests, the proportion of pupils misclassified still remained at an estimated 23%; almost a quarter of pupils.

The publication of end of Key Stage results, and the uses to which they were put, received much criticism for their unfairness (Goldstein, 2001). Fitz-Gibbon (1995 and 1997) and Tymms (1997a) conducted a feasibility study and made recommendations to the government for how a national value-added system, which reported the progress made by pupils rather than output measures alone, could be implemented. A value-added system was argued to present a fairer view of pupil progress and school performance, and in 2003, the Department for Education started to publish value-added measures of performance of schools. Tymms and Dean (2004) raised issues with the way in which value-added scores at the end of Key Stage 2 were published. For example, they argued that the small size of many primary schools would result in large fluctuations in their results from year to year and so even when a school's provision remained stable over time, the errors of measurement associated with small samples of children would give rise to false impressions of change. They identified validity issues with the end of Key Stage 1 data; value-added scores were assigned to the primary school where the pupil took the end of Key Stage 2 test but many pupils would have moved schools in between Key Stages 1 and 2, and therefore their progress could not be attributed to just one school. Since their introduction, the value-added models have become more sophisticated to take account of contextual factors such as entitlement to free school meals.

The Early Years Foundation Stage Profile was introduced in 2003. It was welcomed by some early years professionals but questions exist about its reliability and validity since official statistics of these properties have not been published. There was a requirement to assess all children at the end of the early years foundation Stage and to report the information to parents, forward it to the next teacher and submit it for central collation. The profile was completed on the basis of practitioners' judgements based on their observations of children's behaviour and actions within the school setting and information from parents.

To make comparisons against other countries, England participates in international studies (PIRLS; Progress in International Reading Study, PISA; Programme for International Student Assessment, and TIMSS; Trends in International Maths and Science Survey).

The Northern Ireland System

Northern Ireland's Executive Department of Education (DENI) is responsible for the country's compulsory education policy and the statutory assessments fall within the remit of the Council for the Curriculum, Examinations and Assessment (CCEA).

Education in Northern Ireland begins with the Foundation Phase for children aged 4 to 6 years. Key Stage 1 covers ages 6 – 8 years; Key Stage 2 covers ages 8 –

11 years; Key Stage 3 covers ages 11 – 14 years; and Key Stage 4 covers ages 11 – 15 years. Pupils then study for their end of compulsory education qualifications (GCSE; General Certificate of Education), which are taken at age 16. In contrast to Scotland, Wales and the vast majority of England, Northern Ireland retained a selective secondary education system and, until recently, children were tested at age 11 for selection into academically elite grammar schools or secondary schools. This selection system fell from favour, politically, and 2008 was the last year in which the ‘transfer test’, as it was known, was officially conducted for entry to secondary school in 2009. The minister proposed that from that date, secondary schools should select their intake on the basis of non-academic criteria but to facilitate transition, secondary schools could select up to 50% of their intake on the basis of academic ability for the 2010 intake and then the proportion subsequently reducing. By 2013, it was intended that all secondary schools were to select their intake on the basis of non-academic criteria and therefore no transfer test would be provided by the Department (DENI, 2013). However, despite this recommendation, some schools have continued to select pupils using unregulated tests.

A further policy was launched in 2009; ‘Every School a Good School’ (Department of Education, 2009), which set out expectations for schools to set their own targets for pupils’ literacy and numeracy development, and to monitor progress effectively. In subsequent years, Northern Ireland has continued to ‘embrace the principles of assessment for learning by placing formative assessment at the heart of the learning and teaching cycle’ (Northern Ireland Curriculum, 2013). Teacher assessment is currently used to monitor pupils’ progress from the Foundation Phase to the end of compulsory education when pupils sit examinations set by a range of awarding bodies, one of which is CCEA. New assessment arrangements were introduced in September 2012; a statutory requirement for teachers to assess the cross-curriculum areas of ‘communication’ and ‘using mathematics’ at the end of Key Stages 1, 2 and 3. An assessment of ‘using information and communication technology (ICT)’ will become compulsory from September 2013. Teachers judge the level of each pupil on the basis of their observations supported by regular in-school assessments, and the levels of attainment for all pupils are collated centrally by CCEA (CCEA; 2012a, 2012b) for analysis at school and district level, and the further exploration of other demographic groups.

Northern Ireland participates in PIRLS, PISA and TIMSS.

The Scottish System

In 2010, a new curriculum; the ‘Curriculum for Excellence’, was introduced into Scottish schools for learners aged between 3 and 18 years with the aim of equipping them with high levels of achievement that would enable them to succeed in the workplace in the 21st Century. This was the culmination of a long process which was launched in 2004 (Scottish Executive, 2004). The Curriculum for Excellence provides high-level guidance but it is expected that the detail of the curriculum is developed locally in local authorities in collaboration with schools

rather than imposing a prescriptive approach. Scotland's decentralised approach has been described by Ellis (2007) as putting education into the 'hands of the practitioners'. The guidance comprises a series of experiences and outcomes which are presented in a developmental order for various curricular categories with literacy and numeracy being cross-curricular. The curriculum guidance was accompanied by a framework for assessment (Scottish Government, 2010). Prior to the publication of this framework in 2010, there was some uncertainty about what format a national assessment system would take. There was an appreciation of the danger of assessment driving the curriculum, with advice that the experiences and outcomes were not designed to be assessment criteria in their own right, but this was contradicted to some extent by the suggestion that the experiences and outcomes should allow for the evaluation of pupils' progress (Priestly and Humes, 2010). The assessment framework built upon the existing focus of assessment in Scotland; 'Assessment is For Learning (AiFL), an approach which was underpinned by the research of Black and Wiliam (1998). Black and Wiliam proposed that the wealth of information about pupils' learning, progress and difficulties could be used by both teachers and the pupils themselves to inform subsequent learning, i.e. for formative purposes. They supported active engagement in the assessment process by the learners in order to achieve higher educational outcomes. Building upon this established method, the assessment framework for the Curriculum for Excellence advised teachers to use a range of approaches to assess the "breadth, challenge and application of learning and the wide range of skills being developed" (Scottish Government, 2010). Exemplars were made available via the National Assessment Resource to enable teachers to benchmark their own judgements against agreed standards.

The Scottish Government does not currently collect information on all pupils through national assessments to monitor progress and standards at a system level. However, it does expect schools to be able to report information about improvements in their practices that have led to improvements in pupils' outcomes. Education authorities are expected to have moderated their schools' assessment outcomes against national benchmarks and to be able to feed information into the National Performance Framework.

The Scottish Survey of Literacy and Numeracy is a sample survey which is currently used to monitor standards over time. This assesses pupils at ages 8, 11 and 13 years. Scotland also participates in international studies (PIRLS, PISA and TIMSS) in order to be able to compare the standards of attainment of its pupils against those from other countries.

The Welsh System

The Welsh Board for Education was created in 1907 although decisions for educational policy did not begin to be devolved to Wales until 1999. The 2002 Education Act finally enabled decisions on the school curriculum and assessment to be made by the Welsh Assembly. Today, the provision of education and

UNDERSTANDING MONITORING IN THE UNITED KINGDOM CONTEXT

assessment of pupils' progress in maintained schools is managed by the Department for Education and Skills.

The Stages of education in Wales begin with the Foundation Phase for pupils aged 3 to 7 years, with compulsory education starting at age 5. The Foundation Phase is a relatively new initiative which began in 2008 and the final phase of implementation when the first cohort of pupils reached their fourth year at age 7 began in August 2011 (Welsh Statutory Instruments, 2008). Statutory assessment at the end of the Foundation Phase is through teacher assessment based on observations of children's everyday classroom activities (Welsh Assembly Government, 2011). Following on from the Foundation Phase, compulsory education in Wales is divided into Key Stage 2 (for pupils aged 7 to 11 years), Key Stage 3 (for pupils aged 11 to 14 years) and Key Stage 4 (for pupils aged 14 to 16). Between 1992 and 2001, school performance tables were published but these were abolished for all age-groups in 2001 and in 2004, the Welsh National Assembly abandoned statutory tests at the end of Key Stages 2 and 3; a decision taken after Daugherty Review of assessment policy Key Stages 2 and 3 in Wales (Daugherty, 2004). The Review group heard evidence that many pupils aged 11 and 14 experienced an excessive amount of test preparation and practice, which had led to a narrowing of the curriculum. The Review strongly advised a reduction in the inappropriate use of attainment data and this included the practice of setting targets for cohorts of pupils without taking account of the prior attainment of those particular pupils, and using statutory assessment data without reference to other indicators to evaluate the performance of teachers, schools and districts. It stated that statutory assessment data, when used inappropriately, had the potential to have a negative effect on educational provision. Moreover, it recommended that statutory teacher assessment should be used to provide data on pupil attainment at the end of the Key Stages and that the statutory end of Key Stage tests should be phased out. Daugherty was mindful of the need for the data arising from the teacher assessments to be reliable and acknowledged the professional development needed to achieve that, suggesting a timescale for implementing moderation procedures. Another recommendation arising from the Review was for Assessment for Learning to be embedded across the education system. To be able to compare the performance of pupils at a system level with other countries, the Daugherty Review (2004) recommended that Wales should participate in PISA (The OECD's Programme for International Assessment) from 2006 onwards. Statutory assessment in the foundation phase and Key Stages 2 and 3 in Welsh schools is currently wholly based on teacher assessment; there are no statutory tests (Department for Education and Skills, 2013). At the end of Key Stage 4, pupils sit the General Certificate of Secondary Education (GCSE) examinations, which are set and marked by awarding bodies (NFER, 2011).

Almost a decade after these momentous changes to assessment and monitoring in the Welsh education system, there is a focus on whether or not they have been successful. In 2010, the Welsh school inspectorate (Estyn, 2010) reported that assessment was one of the weakest areas of work in schools and that teachers were not making use of the comprehensive guidance available to them from the

CHRISTINE MERRELL

Department. Estyn found the assessment outcomes in Key Stage 3 to be more reliable than those from Key Stage 2. They also highlighted weaknesses in the way that teachers in secondary schools made use of the data from their associated primary schools, which was perhaps not surprising if the reliability of the data from the end of Key Stage 2 was questionable. A study by Collins, Reiss and Stobart (2010), which was conducted four years after the abolition of statutory tests at the end of Key Stage 2, found that teachers of pupils at the end of Key Stage 2 supported the change from tests to teacher assessment. The teachers were specifically asked about the changes in relation to science and they identified positive implications for academic and attitudinal domains. Many reported greater flexibility to determine the content of lessons whilst nevertheless working within the National Curriculum. The use of teacher assessment rather than tests had meant that they changed their teaching strategies so that they were able to closely observe pupils and that had, in turn, led to providing experiences that were more closely aligned to the needs of individual pupils and to include more practical activities than previously. However, not all those teachers surveyed favoured the changes. A quarter of the teachers and head teachers surveyed reported lacking confidence in teacher assessment to provide reliable judgements of pupils' science attainment, and many used optional tests to validate their judgements.

ARE THE CURRENT UK ASSESSMENT & MONITORING POLICIES ROBUST AND USEFUL?

From the overview of the recent developments in the assessment and monitoring of pupils' progress in the four countries within the UK, a shift from national tests towards teacher assessment is evident, although less so in England. In Northern Ireland, Scotland and Wales, there is a strong emphasis on teacher assessment for formative purposes; using the information from assessments for planning next steps and a close monitoring of progress. England has currently retained some tests at the end of Key Stage 2 but other statutory assessments are based on teachers' judgements.

Whilst teacher assessment has some advantages, there are also potential problems, some of which have been noted in the previous section, in particular within the discussion of the Welsh system. Harlen (2004 & 2005) systematically reviewed the evidence of reliability and validity of teacher assessment for summative purpose and found instances of bias in teacher assessments in relation to factors including sex, ability, ethnicity, social class, age and behaviour.

The previous sections of the chapter have given an overview of some of the uses of the statutory data in each country. England differs from the other UK countries in the way that it continues to publish school performance tables. Some of the negative effects of this policy on pupils and teachers have been described but what effect might the use of data for accountability have had on standards over time? Have the standards of pupils' outcomes improved as a result of the methods of monitoring? It is difficult to tell. One way is to look at each country's performance

and ranking in the international studies of PISA, TIMSS and PIRLS, but there are other studies too, some of which are discussed below.

In England, the statutory test data from the end of Key Stage 2 (age 11) was reported as increasing steadily and at an unexpectedly rapid rate between 1995 and 2000. Tymms (2004) investigated this large rise by comparing the scores from the statutory tests with data from the other independent studies that had collected data over the same period. Overall, data from the independent studies consistently showed a rise in scores between 1995 and 2000 but this improvement was much smaller than the statutory test data. The huge number of educational initiatives introduced in England over that period, including the publication of school-level results in league tables, did not appear to be associated with significant improvements in pupils' outcomes at the end of primary school.

Burgess, Wilson and Worth (2010) used the abolition of published school performance tables in Wales to analyse the results of what they described as a 'natural experiment' to investigate differences in school effectiveness between England and Wales. Wales published tables of secondary school performance, as measured by pupil performance in the General Certificate of Secondary Education examinations taken at the end of compulsory education, from 1992 up to 2001. England also published league-tables of the same information and continues to do so. Burgess et al. tested the hypothesis that school effectiveness in Wales, after the abolition of league tables, would be lower than England. They suggested that the league tables were scrutinized by parents who may then take action to avoid sending their children to low-performing schools, and by education authorities who may impose sanctions on low-performing schools. The data were also used by the school inspection systems as an element of their judgement of schools. Burgess et al. compared the results of the pupils in the cohort which took their GCSE examinations in 2001 with cohorts from later years and cohorts in England over the same period. They found that the reforms in Wales significantly reduced average performance and increased educational inequality. The authors did acknowledge that their analysis did not take account of previously-reported negative effects of publishing results, including teacher and head-teacher morale, and narrowing of the curriculum. Despite this, their findings offered an interesting perspective for consideration alongside the widely discussed negative aspects of the publication of school performance tables. Their final concluding remarks were that school accountability policies hold promise for raising school performance and that making schools' results public appeared to be one cost-effective method of accountability. In a more recent publication by Burgess (2013), he expanded upon this study by suggesting the type of information that would be useful to publish about school performance so as to further enhance schools' effectiveness whilst reducing some of the negative associations.

By contrast to the findings from Burgess et al. (2010), which indicated the emergence of a difference between school performance in England and Wales at the same time as there was a change in policy, (Machin, McNally and Wyness, 2013) found that all four countries (England, Northern Ireland, Scotland and Wales) attained similar positions relative to the international community in

international studies and they suggested that continued devolution should not, in theory, result in large changes in outcomes. (Machin et al., 2013)

The statutory requirements for the assessment and monitoring of pupil and school performance across the UK are evolving but the results of those assessments, in whatever their form, are still expected to serve a wide range of purposes. Mansell et al. (2009) cautioned that ‘politicians, policy-makers, educators, parents and the general public should be alert to the intended and unintended consequences of assessment policy decisions and should ask whether the policies are truly fit for purpose’. They called for a need to extend best practice, to provide professional development for teachers, the dependability of assessment results to be enhanced and the creation of more intelligent accountability systems.

AN ALTERNATIVE NON-STATUTORY MONITORING SYSTEM FOR SELF-EVALUATION

For thirty years, the Centre for Evaluation and Monitoring (CEM) at Durham University, England, has run school monitoring systems on a large scale (see www.cem.org). These systems cover the 3 – 18 age range and schools, districts and, occasionally, jurisdictions, pay to use them. At the time of writing, approximately one million students are being assessed with CEM’s monitoring systems each year and schools in over 70 countries make use of them. There are large samples of schools in Australia, Abu Dhabi, England and Scotland; smaller samples in Germany, New Zealand and South Africa; and individual international schools registered around the world. Carol Taylor Fitz-Gibbon and Peter Tymms were the main founders of these monitoring systems (See the following references for further description of the systems and the rationale underpinning them: Fitz-Gibbon, 1996; Tymms, 1999; Tymms and Albone, 2002).

The aim of CEM’s monitoring systems is to provide high-quality information about pupils’ attainment, progress, developed ability and attitudes to learning for use by teachers and head teachers. Educators can use the information to identify problems with individuals or groups, and also to identify where things are going well, and tailor provision accordingly. In this sense, the systems are seen as a tool for professional monitoring rather than for public accountability (Tymms, 1998). CEM has developed its own assessments, the majority of which are now computer-adaptive, which teachers administer. Computer-adaptive assessments have an advantage over more traditional methods of administration, being tailored to each pupil’s ability. These assessments are used by all schools registered for CEM’s systems and form a consistent comparison rather than using scores from teachers’ tests, which would not be comparable. Where available and reliable, scores from national statutory assessments (e.g. GCSE scores) are also collected for the older pupils and fed into the analyses. Feedback is generated rapidly and returned to schools. This includes scores for different areas assessed that are generally standardised to enable comparisons against group norms to be made, comparison of the different constructs assessed and examination of changes over time. Predictions of later outcomes are given that give the likelihood of attaining a particular outcome rather than reporting the most likely outcome only. Finally, a

measure of value-added is provided (which is calculated using ordinary least squares regression) so that teachers and head teachers can see whether their pupils are making expected progress in relation to their previous attainment or against their developed ability (this is derived from curriculum-free measures of pupils' vocabulary acquisition, non-verbal ability in the primary years, and vocabulary, non-verbal ability, mathematical reasoning and other curriculum-free skills in the secondary years). Developed ability offers a further dimension when evaluating a pupil's progress. Simply looking at a pupil's performance in, for example, mathematics over time will show whether that pupil is making expected progress or not compared with other pupils. However, it is possible that although good progress is being made, a pupil is nevertheless underachieving given his/her developed ability, which is useful information for teachers.

Assessment results need to be easily understood by teachers and others who are not necessarily experts in understanding assessment data and statistical concepts. To this end, CEM's researchers have found ways of presenting complex information in easily-understood formats. Tymms (1997b) randomly assigned different formats of feedback to a sample of schools to investigate which was most readily understood; tables or graphs. Participants reported a preference for information presented in tables. Additionally, the pupils in the schools who received the tabular format attained higher scores in their later end of Key Stage 2 statutory tests.

Having received the results from assessments, teachers commonly ask what to do next. In other words, how do they relate the assessment scores to teaching and learning strategies? For some of its systems, CEM provides research-based advice in the form of short booklets for teachers about how to help pupils with particular profiles to improve in the form of short booklets. For example, the Centre has developed a computer-adaptive, diagnostic assessment of reading and mathematics for pupils aged between 6 and 11 years. For reading, this assessment provides a profile of each pupil's performance in word recognition, word decoding and comprehension. It also includes optional assessments of pupils' spelling, vocabulary acquisition and non-verbal ability. This provides a powerful profile of a pupil's strengths and weaknesses. Examples of pupils with, say, good word recognition and decoding ability but poor comprehension skills, are linked to effective research-based strategies within the advice booklets (Merrell and Tymms, 2007).

CEM's monitoring systems have thrived for decades and their use is increasing alongside statutory assessments which are used for public accountability. The assessments themselves are easy to administer, which teachers appreciate, and they are enjoyed by pupils. Another consideration is the speed of processing pupils' scores and returning them to schools. This processing frequently takes just a few seconds, sometimes twenty-four hours, saving teachers hours of marking scripts. Instead, they can assign that time to interpreting the information and using it effectively. These features are lacking in the UK's statutory systems. There has been a recent sharp rise in the number of local authorities and schools in Scotland

CHRISTINE MERRELL

using them, which coincides with the changes associated with the Curriculum for Excellence and its associated assessment system.

The systems go some way towards meeting the criteria suggested by Mansell et al. (2009) in that they provide an intelligent and sophisticated way of assessing and monitoring progress. However, to be able to continue to do so, the data needs to continue to be interpreted and used in an appropriate way.

UNDERSTANDING MONITORING IN THE UNITED KINGDOM CONTEXT

REFERENCES

- Alexander, R. (Editor) (2010), *The Cambridge Primary Review Research Surveys*. Pub. Oxon: Routledge.
- Black, P. & Wiliam, D. (1998) *Inside the Black Box: Raising Standards Through Classroom Assessment*. Pub. London: King's College London School of Education.
- Bew, Lord (2011) *Independent Review of Key Stage 2 testing, assessment and accountability: Final Report*. www.education.gov.uk Accessed 02 July 2011.
- Burgess, S. (2013) *School Accountability, Performance and Pupil Attainment*, In Clifton (Ed) *Excellence & Equity: Tackling Disadvantage in England's Secondary Schools*. Pub. London: Institute for Public Policy Research.
- Burgess, S., Wilson, D. and Worth, J. (2010) *A natural experiment in school accountability: the impact of school performance information on pupil progress and sorting*, Working Paper 10/246. Pub. Bristol: CMPO, Centre for Marketing and Public & Public Organisation, Bristol Institute of Public Affairs. www.bristol.ac.uk/cmppo/publications/papers/2010/wp246.pdf
- CCEA (2012a) *Assessment Arrangements in Relation to Pupils in the Final Years of Key Stages 1 & 2 for Communication and Using Mathematics, 2012 – 13, Revised*. Pub. Belfast: CCEA.
- CCEA (2012b) *Assessment Arrangements in Relation to Pupils in the Final Years of Key Stage 3 for Communication and Using Mathematics, 2012 – 13, Revised*. Pub. Belfast: CCEA.
- Collins, S., Reiss, M. & Stobart, G. (2010) *What happens when high-stakes testing stops? Teachers' perceptions of the impact of compulsory national testing in science of 11-year-olds in England and its abolition in Wales*, *Assessment in Education: Principles, Policy & Practice*, 17:3, 273-286, <http://dx.doi.org/10.1080/0969594X.2010.496205> . Accessed 08/09/13.
- Daugherty, R. (2004) *Learning pathways through statutory assessment: Key Stages 2 and 3. Final Report of the Daugherty Assessment Review Group*, Cardiff: Welsh Assembly Government.
- Davies, J. and Brember, I. (1998) *National curriculum testing and self-esteem in year 2 the first five years: a cross-sectional study*, *Educational Psychology*, 18, pp. 365–375.
- Davies, J. and Brember, I. (1999) *Reading and mathematics attainments and self-esteem in years 2 and 6: an eight year cross-sectional study*, *Educational Studies*, 25, pp. 145–157.
- DENI (2013) *Post-primary transfer policy*. <http://www.deni.gov.uk/post-primary-transfer-policy-from-september-2013-english-pdf-1.76-mb.pdf> Accessed 06/10/13.
- Department of Education (2009) *Every School a Good School: A Policy for School Improvement*. http://www.deni.gov.uk/esags_policy_for_school_improvement_-_final_version_05-05-2009.pdf Accessed 09/10/13.
- Department for Education and Skills (2013) *Statutory Assessment Arrangements for the End of Foundation Stage and Key Stages 2 and 3: Guidance*, Guidance Document No. 096/2013, February 2013. <http://wales.gov.uk/docs/dcells/publications/130219-statutory-assessment-arrangements-2012-13-booklet-en.pdf> . Accessed 08/09/2013.
- Estyn (2010). *Evaluation of the arrangements to assure the consistency of teacher assessment in the core subjects at Key Stage 2 and Key Stage 3*. Cardiff: Estyn. www.estyn.gov.uk . Accessed 08/09/13
- Ellis, S. (2007). *Policy and Research: Lessons from the Clackmannanshire Synthetic Phonics Initiative*. *Journal of Early Childhood Literacy*, 7, 281-293.
- Fitz Gibbon, C.T., (1995), *The Value Added National Project: Issues to be considered in the design of a National Value Added System*, Pub. London: SCAA Publications.
- Fitz-Gibbon, C.T. (1996) *Monitoring Education: Indicators, Quality and Effectiveness*, Pub. London: Cassell.
- Fitz Gibbon C., (1997) *The value added national project, final report, feasibility studies for a national system of value-added indicators*, Pub. London: SCAA publications.
- Goldstein, H. (2001) *Using Pupil Performance Data for Judging Schools and Teachers: Scope and Limitations*, *British Educational Research Journal*, 27(4), p432 – 442.
- Harlen, W. (2004). *A systematic review of the evidence of reliability and validity of assessment by teachers used for summative purposes*. In the *Assessment and Learning Research Synthesis Group*

CHRISTINE MERRELL

- Research Evidence in Education Library. London: EPPI-Centre, Social Science Research Unit, Institute of Education.
- Harlen, W. (2005). Trusting teachers' judgement: research evidence of reliability and validity of teachers' assessment used for summative purposes. *Research Papers in Education*, 20, 245-270.
- James, M. (2013) *Educational Assessment, Evaluation and Research: The Selected Works of Mary James*. Pub. London: Routledge.
- Machin, S., McNally, S. and Wyness, G. (2013) *Education in a Devolved Scotland: A Quantitative Analysis*. Report to the Economic and Social Research Council, May 2013. <http://cep.lse.ac.uk/pubs/download/special/cepsp30.pdf> Accessed 15/10/13.
- Mansell, W., James, M. & the Assessment Reform Group (2009) *Assessment in schools. Fit for purpose? A Commentary by the Teaching and Learning Research Programme*. London: Economic and Social Research Council, Teaching and Learning Research Programme. <http://www.tlrp.org/pub/documents/assessment.pdf> . Accessed 07/09/13
- Merrell, C. and Tymms, P. (2007) Identifying Reading Problems with Computer-Adaptive Assessments, *Journal of Computer Assisted Learning*, 23(1) p27 – 35.
- National Union of Teachers (2006) Briefing: The Impact of National Curriculum Tests on Pupils.
- Northern Ireland Curriculum (2013) http://www.nicurriculum.org.uk/Key_Stages_1_and_2/assessment/ Accessed 06/10/13.
- NFER (2011) Overview of education system in Wales, Eurydice at NFER, November 2011. http://www.nfer.ac.uk/shadomx/apps/fms/fmsdownload.cfm?file_uuid=F0943FAA-C29E-AD4D-02BA-AA5A783AED68&siteName=nfer . Accessed 08/09/13.
- Phillips, R. (2003) Education policy, comprehensive schooling and devolution in the disUnited Kingdom: an historical 'home international' analysis, *Journal of Education Policy*, 18:1, 1-17. <http://dx.doi.org/10.1080/0268093032000042173> . Accessed 07/09/13
- Priestley, M. & Humes, W. (2010) The development of Scotland's Curriculum for Excellence: amnesia and déjà vu, *Oxford Review of Education*, 36:3, 345-361.
- Reay, Y. and Wiliam, D. (1999) 'I'll be a nothing': structure, agency and the construction of identity through assessment, *British Educational Research Journal*, 25, pp. 343-354.
- Scottish Executive (2004) *A Curriculum for Excellence: the curriculum review group*. Pub. Edinburgh: Scottish Executive.
- Scottish Government, The (2010) *Building the Curriculum 5: A Framework for Assessment*. Pub. Edinburgh: The Scottish Government. ISBN: 978-0-7559-8050-5
- Task Group on Assessment and Testing (1988) *National Curriculum Task Group on Assessment and Testing: A Report for the Department of Education and Science and the Welsh Office*. <http://www.educationengland.org.uk/documents/pdfs/1988-TGAT-report.pdf> Accessed 09/10/12.
- Tymms, P. (1997a), *The Value Added National Project: Technical Report: Primary 4 : Value-added Key Stage 1 to Key Stage 2*. Pub. London: SCAA.
- Tymms, P.B. (1997b). Responses of Head teachers to Value-added and the impact of feedback. *Value Added National Project*. London: School Curriculum and Assessment Authority.
- Tymms, P.B. (1998). Accountability and Quality Assurance. In C. Richards & P. Taylor (Eds.), *How shall we school our children?* London: Falmer Press.
- Tymms, P. (1999). Baseline assessment and monitoring in primary schools: achievements, attitudes and value-added indicators. London, David Fulton.
- Tymms, P. (2004) Are Standards Rising in English Primary Schools? *British Educational Research Journal*, 30(4) p477 – 494.
- Tymms, P., & Albone, S. (2002). Performance Indicators in Primary Schools. *School Improvement Through Performance Feedback*. In A. J. Visscher & R. Coe. Lisse/Abingdon/Exton PA/Tokyo, Swetz & Zeitlinger (p191-218).
- Tymms, P. & Dean, C. (2004) *Value-Added in the Primary School League Tables: A Report for the National Association of Head Teachers*. Pub. Durham: Centre for Evaluation & Monitoring, Durham University.
- Welsh Assembly Government (2011) *End of Foundation Phase Assessment: Exemplification of Outcomes*. Cardiff: Department for Children, Education, Lifelong Learning and Skills. <http://wales.gov.uk/docs/dcells/publications/110323endofphaseen.pdf> Accessed 04/10/2013.

UNDERSTANDING MONITORING IN THE UNITED KINGDOM CONTEXT

- Welsh Statutory Instruments (2002) An Act to Make Provision about Education, Training and Childcare. http://www.legislation.gov.uk/ukpga/2002/32/pdfs/ukpga_20020032_en.pdf . Accessed on 08/09/13
- Welsh Statutory Instruments (2008) The Education (National Curriculum) (Foundation Stage) (Wales) Order 2008, No. 1732 (W.169). <http://www.legislation.gov.uk/wsi/2008/1732/introduction/made> . Accessed 08/09/13
- William, D. (2000) Reliability, validity and all that jazz, Education 3 - 13, 29(3) 9 – 13.
- Whetton, C. (2009) A brief history of a testing time: national curriculum assessment in England 1989-2008, Educational Research, 51(2): 137-159. <http://dx.doi.org/10.1080/00131880902891222> . Accessed 07/09/13

Christine Merrell
CEM,
Durham University